Physics Letters B 774 (2017) 564–568

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Machine-learning the string landscape

Yang-Hui He [a,b,c,*]

[a] Merton College, University of Oxford, UK
[b] Department of Mathematics, City, University of London, UK
[c] School of Physics, NanKai University, China

ARTICLE INFO

Article history:
Received 5 September 2017
Received in revised form 4 October 2017
Accepted 10 October 2017
Available online 16 October 2017
Editor: N. Lambert

ABSTRACT

We propose a paradigm to apply machine learning various databases which have emerged in the study of the string landscape. In particular, we establish neural networks as both classifiers and predictors and train them with a host of available data ranging from Calabi–Yau manifolds and vector bundles, to quiver representations for gauge theories, using a novel framework of recasting geometrical and physical data as pixelated images. We find that even a relatively simple neural network can learn many significant quantities to astounding accuracy in a matter of minutes and can also predict hithertofore unencountered results, whereby rendering the paradigm a valuable tool in physics as well as pure mathematics.

1. Introduction

Whereas theoretical physics now inevitably resides in an Age where new physics, new mathematics and new data coexist in a symbiosis transcending disciplines, string theory has spearheaded this vision. That it engenders the cross-fertilization between physics and pure mathematics is without dispute, that it also has been a testing ground for computational mathematics and "big data" is perhaps less known. With the advent of increasingly powerful computers, from this fruitful dialogue has also arisen a plethora of data, ripe for mathematical experimentation. This emergence of data began with the incipience of string phenomenology [1] where compactification of the heterotic string on Calabi–Yau threefolds (CY3) was widely believed to hold the ultimate geometric unification. A race, spanning the 1990s, to explicitly construct examples of Calabi–Yau (CY) manifolds ensued, beginning with the so-called complete intersection CY manifolds (CICYs) [2], proceeding to the hypersurfaces in weighted projective space [3], to elliptic fibrations [4] and ultimately culminating in the impressive (at least some 10^{10}) list of CY3s from reflexive polytopes [5].

With the realization that the landscape of stringy vacua might in fact exceed the number of inequivalent CY3s [6] by hundreds of orders of magnitude, there was a vering of direction toward a more multi-verse or anthropic philosophy. Nevertheless, hints have emerged that the vastness of the landscape might well be mostly infertile (cf. the swamp-land of [7]) and that we could live in a very special universe [8–10], a "des res" corner within a barren vista.

Thus, undaunted by the seeming over-abundance of possible vacua, fortified by the rapid growth of computing power and inspirited by the omnipresence of big data, the first two decades of the new millennium saw a return to the earlier principle of creating and mining geometrical data; the notable fruits of this combined effort between pure and computational algebraic geometers as well as formally and phenomenologically inclined physicists have included (q.v. [11] for a review of the various databases): (1) Continuing with Kreuzer–Skarke (KS) database [14–22]; (2) Generalizing the CICY construction [23–28]; (3) Finding elliptic and K3 fibred CY for F-theory and string dualities [13, 14,28–32]; (4) D-brane world-volume theories as supersymmetric quiver gauge theories [33–41].

All of the above cases are accompanied by typically accessible data of considerable size, representing a concrete glimpse onto the string landscape, to which we shall refer as **landscape data**. For instance, the heterotic line bundles on CICYs are on the order of 10^{10}, the spectral-cover bundles on the elliptically fibred CY3, 10^6, the brane-configurations in the CY volume studies, 10^5, type II intersecting brane models, 10^9, etc. Even by today's measure, these constitute a fertile playground of data, the likes of which Google and IBM are constantly analysing. A natural course of action, therefore, is to do unto this landscape data, what Google et al. do each second of our lives: to machine-learn.

Let us be precise about what we mean by *deep machine-learning* this landscape. Much of the aforementioned data have been the

* Correspondence to: Merton College, University of Oxford, UK.
E-mail address: hey@maths.ox.ac.uk.

https://doi.org/10.1016/j.physletb.2017.10.024

brain-child of the marriage between physicists and mathematicians, especially incarnated by applications of computational algebraic geometry, numerical algebraic geometry and combinatorial geometry to problems which arise from the classification in the physics and recast into a finite, algorithmic problem in the mathematics (cf. [12]). Obviously, computing power is a crucial limitation. Unfortunately, in computational algebraic geometry – on which most of the data heavily rely, ranging from bundles stability in heterotic compactification to Hilbert series in brane gauge theories – a decisive step is finding a Groebner basis, which is notoriously known to be unparallelizable and double-exponential in running time. Thus, much of the challenge in establishing the landscape data had been to either circumvent the direct calculation of the Groebner bases by harnessing of the geometric configuration – e.g., using the combinatorics when dealing with toric varieties. Still, many of the combinatorial calculations, be they triangulation of polytopes or finding dual cones, are still exponentially expensive.

The good news for our present purpose is that, *much of the data have already been collected*. Oftentimes, as we shall find out in our forthcoming case-studies, tremendous effort is needed for deceptively simple questions. Hence, to draw inferences from *actual* theoretical data by deep-learning therefrom would not only help identify undiscovered patterns but also aid in predicting results which would otherwise cost formidable computations. Subsequently, we propose our

> **Paradigm:** To set-up neural networks (NN) to deep-learn the landscape data, to recognize unforeseeable patterns (as classifiers) and to extrapolate to new results (as predictors).

Of course, this paradigm is useful not only to physicists but to also to mathematicians; for instance, could our NN be trained well enough to approximate bundle cohomology calculations? This, and a host of other examples, we will now examine.

Methodology Neural networks are known for their complexity, involving usually a complicated directed graph each node of which is a "perceptron" (an activation function imitating a neuron) and amongst the multitude of which there are many arrows encoding input/output. Throughout this letter, we will use a rather simple multi-layer perceptron (MLP) consisting of 5 layers, three of which are hidden, with activation functions typically of the form of a logistic sigmoid or a hyperbolic tangent. The input layer is a linear layer of 100 to 1000 nodes, recognizing a tensor (as we will soon see, algebro-geometric objects such as Calabi–Yau manifolds or polytopes are generically configurations of integer tensors) and the output layer is a summation layer giving a number corresponding to a Hodge number, or to rank of a cohomology group, etc. Such an MLP can be implemented, for instance, on the latest versions of Wolfram Mathematica. With 500–1000 training rounds, the running time is merely about 5–20 minutes on an ordinary laptop. It is reassuring and pleasantly surprising that even such a relatively simple NN can achieve the level of accuracy shortly to be presented.

This letter is a companion summary of the longer paper [42] where the interested reader can find more details of the computations and the data.

2. Results

With simple NNs, we proceed to analyse our landscape data, a fertile ground constituting more than 2 decades of many international collaborations between physics and mathematicians. Using 4 concrete case studies, we first "learn" from the inherent structure and then "predict" unseen properties; considering how difficult some of the calculations involved had been in establishing the databases, the usefulness of our paradigm is evident.

2.1. Case study 1: CY hypersurfaces in $W\mathbb{P}^4$

One of the first datasets [3] to experimentally illustrate mirror symmetry was that of hypersurfaces in weighted projective space $W\mathbb{P}^4$. The ambient space $W\mathbb{P}^4_{[w_0:w_1:w_2:w_3:w_4]}$ with weights $w_{i=0,\ldots,4} \in \mathbb{Z}_+$ is in general singular, but a generic enough homogeneous polynomial of degree $\sum_{i=0}^{4} w_i$ which misses the singularities defines a hypersurface therein which is a smooth CY3 X. There are 7555 inequivalent such configurations, each specified by a 5-vector $\vec{w}_{i=0,\ldots,4}$. The Euler characteristic χ of X is easily given in terms of the vector. However, as is usually the case, the individual Hodge numbers $(h^{1,1}, h^{2,1})$ are less amenable to a simple combinatorial formula. The original computation resorted to Landau–Ginzberg techniques to obtain the list of Hodge numbers [3]. One could in principle use adjunction and Euler sequences, and singularity resolution, but this is not an easy task to automate.

Suppose we have a simple question: *how many such CY3s have a relatively large number of complex deformations*? We can, for instance, consider $h^{2,1} > 50$ to be "large" and let training data be of the form $w_i \to 1$ or 0 depending on whether $h^{2,1}(X) > 50$. Training the NN, with say 500 rounds, takes under a minute on an ordinary laptop. The result is an optimised continuous real output between 0 and 1, the rounding of which can then be compared with the actual data. An accuracy of 96.2 % is achieved almost effortlessly! To appreciate the *predictive* power of the network, suppose that we only had partial data. This is particularly relevant when for instance, due to computational limitations, a classification is not yet complete, or when a quantity in question has not been or could not be yet computed.

Therefore, let us pretend that we have only data available for the first 3000 out of the 7555 $(X, h^{2,1})$ pairs. We repeat the procedure on the 3000, and then test against the full 7555. We find that 6078 cases were actually correct. Thus, with rather incomplete training data, the NN has learnt, in under a minute, our question and predicted new results to 80% accuracy.

Emboldened, let us move onto another question, of importance to string phenomenology: *Given a configuration, can one tell whether χ is a multiple of 3?* In the early days of heterotic string compactification, this question was decisive on whether the model admitted 3 generation of particles in the low-energy effective gauge theory. Again, we can define a binary function taking the value of 1 if $\chi \bmod 3 \equiv 0$ and 0 otherwise. Training with the NN, we achieve 82% accuracy with 1000 training rounds, taking about 2 minutes; these figures are certainly expected to improve with increasing number of training rounds and with more layers or more nodes in the NN.

The astute reader might question at this stage why we have adhered to *binary queries*. Why not train the NN to answer a direct query, i.e., to try for instance to learn and predict the value of $h^{2,1}$ itself? This is a matter of spread in the present dataset: we have only some 10^4 inputs yet we can see that the values of $h^{1,1}$ ranges from 1 to almost 500. We do not have enough data here to make more accurate statements. This is precisely in line with our philosophy, the power of deep-learning the landscape lies in rapid *estimates*, in identifying patterns and drawing inferences and in avoiding intense computations.

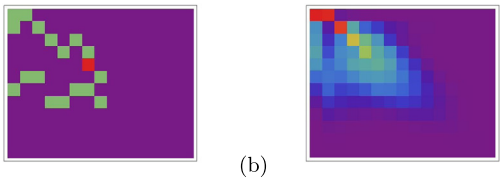

Fig. 1. We realize the set of 7890 CICYs (Calabi–Yau threefolds as complete intersections in products of projective spaces) as 12×15 matrices, padding with zeros where necessary. Then all CICY configurations are such matrices with entries in $\{0, 1, 2, 3, 4, 5\}$. We consider these as pixel colours and draw a typical CICY in (a), with 0 being purple. In (b), we average over all such matrices component-wise, and draw the "average" CICY as a pixelated image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Case study 2: CICYs

Having warmed up, let us move onto complete intersection Calabi–Yau threefolds (CICYs) in products of projective spaces. This is both the first Calabi–Yau database (or, for that matter, the first database in algebraic geometry) [2] and the most heavily studied recently for string phenomenology [23–26,28]. It has the obvious advantage that the ambient space is smooth by choice.

Briefly, CICYs embed as K homogeneous polynomials in $\mathbb{P}^{n_1} \times \ldots \times \mathbb{P}^{n_m}$, of multi-degree q_j^r, with complete intersection meaning that $K = \sum_{r=1}^{m} n_r - 3$ and CY condition implying $\sum_{j=1}^{K} q_j^r = n_r + 1 \ \forall r = 1, \ldots, m$. The construction of CICYs is thus reduced to a combinatorial problem of classifying the integer matrices. The most famous CICY is, of course, [4|5] or simply the matrix [5], denoting the quintic hypersurface in $\mathbb{P}^4$. It was shown that such configurations are finite in number and the best available computer at the time (1990's), viz., the super-computer at CERN [2], was employed. A total of 7890 inequivalent manifolds were found, corresponding to matrices with entries $q_j^r \in [0, 5]$, of size ranging from 1×1 to maximum number of rows and columns being 12 and 15, respectively.

This representation is much in the standard way to represent an image: to pixelate it into blocks of $m \times n$, each of which carrying a colour info, for example, a 3-vector encapturing the RGB data. Therefore, we can represent all the 7890 CICYs into 12×15 matrices over $\mathbb{Z}/6\mathbb{Z}$, embedded starting from the upper-left corner, say, and padding with zeros everywhere else, as illustrated in Fig. 1. To view a CICY as a pixelated image, and indeed, to use image processing to address problems in geometry and mathematical physics, is an entirely new idea worthy of extensive exploration.

Can we deep-learn, say the full list of Hodge numbers? As usual, the Euler number is relatively easy to obtain and there is a combinatorial formula in terms of the integers q_j^r, whilst the individual Hodge numbers $(h^{1,1}, h^{2,1})$ involve some non-trivial adjunction and sequence-chasing, which luckily had been performed for us [2]. Again, we set up a list of training rules (padded configuration matrix $\to h^{1,1}$) and find that the NN can be trained to an accuracy of 99.91% in under 10 minutes! What about the NN as a predictor, which is obviously a more salient question? Suppose the NN were trained with the first 5000 of the data, then, checking against the full dataset comprising of configurations/images the NN has never before seen, we achieve 77% accuracy. Considering (1) that we have only trained the NN for a mere 6 minutes, (2) that it has seen only a little over half of the data, (3) that it is rather elementary with only 5 forward layers, and (4) that the variation of the output is integral ranging from 0 to 19, with no room for continuous tuning, such accuracy with so little effort is quite amazing.

2.3. Case study 3: bundle cohomology

The subject of vector bundle cohomology has, since the so-named "generalized embedding" [1] of heterotic compactification on smooth CY3 X endowed with a (poly-)stable holomorphic vector bundle V, become one of the most active dialogues between algebraic geometry and theoretical physics. The realization [9] that the theoretical possibility of [1] can be concretely achieved by a judicious choice of (X, V) to give the exact MSSM spectrum induced much activity in establishing relatively large datasets to see how often this might occur statistically [11,18,22,24,31], culminating in [25,26] which found some 200 out of a scan of 10^{10} bundles which have exact MSSM content.

Upon this vast landscape let us take an insightful glimpse by taking the dataset of [31], which are $SU(n)$ vector bundles V on elliptically fibred CY3. By virtue of a spectral-cover construction [4,30], these bundles are guaranteed to be stable and hence preserves $\mathcal{N} = 1$ supersymmetry in the low-effective action, together with GUT gauge groups E_6, $SO(10)$ and $SU(5)$ respectively for $n = 3, 4, 5$. We take the base of the elliptic fibration – of which there is a finite list [29] – as the r-th Hirzebruch surface ($r = 0, 1, \ldots, 12$ denoting the inequivalent ways which $\mathbb{P}^1$ can itself fibre over $\mathbb{P}^1$ to give a complex surface), in which case the stable $SU(n)$ bundle is described by 5 numbers (r, n, a, b, λ), with $(a, b) \in \mathbb{Z}_+$ and $\lambda \in \mathbb{Z}/2$ being coefficients which specify the bundle via the spectral cover. This ordered 5-vector will constitute our neural input. The database of viable models was set up in [31], viable meaning that the bundle-cohomology groups of V are such that $h^0(X, V) = h^3(X, V) = 0$ and $\left|h^1(X, V) - h^2(X, V)\right| \equiv 0 \ (\mathrm{mod}3)$, where the first is a necessary condition for stability and the second, that the GUT theory has the potential to allow for 3 net generations of particles upon breaking to MSSM by Wilson lines. Over all the Hirzebruch-based CY3, 14,264 models were found; a sizeable play-ground.

Suppose the output be a 2-vector, indicating (I) what the gauge group is, as denoted by n, and (II) whether there are more generations than anti-generations, as denoted by the sign of the difference $h^1(X, V) - h^2(X, V)$; this is clearly a phenomenologically interesting question. With 1000 training rounds on a NN with an output linear layer, and with the dataset consisting of entries in the form $(r, n, a, b, \lambda) \to (n, \mathrm{Sign}(h^1(X, V) - h^2(X, V)))$, in about 10 minutes, we achieve 100% accuracy (i.e., the neural network has completely learnt the data). Training with partial data, say 8000 points, a little over half, achieves 68.9% predicative accuracy over the entire set.

2.4. Case study 4: quiver gauge theories

As a final example let us tackle affine varieties in the context of quiver representations. Physically, these correspond to world-volume gauge theories coming from D-brane probes of geometric singularities in string theory, as well as the space of vacua for classes of supersymmetric gauge theories in various dimension; they have been data-mined since the early days of AdS/CFT (cf. [33,34]). When the geometry concerned is an affine toric CY variety, the realization of brane-tiling [35] has become the correct way to understand the gauge theory and since then databases have begun to be compiled [36,37].

The input data consists of a quiver (directed graph) and a relation imposed by a polynomial super-potential (q.v. [11] for a rapid review). We can succinctly encode the above information into two matrices, which again can be turned into a pixelated image: (1) D-term matrix Q_D, which comes from the kernel of the incidence matrix d of the quiver, each column of which corresponds to an arrow with -1 as head and $+1$ as tail and 0 otherwise;

(2) F-term matrix Q_F each column of which documents where and with what exponent the field corresponding to the arrow appears in ∂W. Concatenating Q_D and Q_F gives the so-call total charge matrix Q_t of the moduli space as a toric variety (q.v. §2 of [34] for the precise procedure). The combinatorics and geometry of the above is a long story spanning a lustrum of research to uncover followed by a decade of still-ongoing investigations.

In the first database of [36], a host of examples were tabulated. A total of 375 quiver theories much like the above were catalogued (a catalogue which has recently been vastly expanded in [37]). Though not very large, this gives us a playground to test some of our ideas. The input data is the total charge matrix Q_t, the maximal of whose number of rows and columns are, respectively 33 and 36, and all taking values in $\{-3, -2, \ldots, 3, 4\}$. Now, suppose we wish to know the number of points of the toric diagram associated to the moduli space, which is clearly an important quantity. In principle, this can be computed (albeit computationally intensive): the integer kernel of Q_t should give a matrix whose columns are the coordinates of the toric diagram, with multiplicity (associated to the perfect matchings of the bipartite tiling). Training with our NN with the full list achieves, in under 5 minutes, 99.5% accuracy.

2.5. A sanity check

Lest the readers' optimism be elevated to unreasonable heights by the string of successes with the NNs, it is imperative that we be aware of deep-learning's limitations. We therefore finish with a sanity check that a NN is not some omnipotent oracle capable of predicting *any* pattern. An example which must be doomed to failure is the primes (or, for that matter, the zeros of the Riemann zeta function). Indeed, if an NN could learn some unexpected pattern in the primes, this would be a rather frightening prospect for mathematics. We test the sequence of primes (i.e., data of the form `i -> Prime[i]`) with our NN, and achieve no better than a 0.1% accuracy. Our NN is utterly useless against this formidable challenge; we are better off trying a simple regression against some $n\log(n)$ curve, as dictated by the prime number theorem. This is a sobering exercise as well as a further justification of the various case studied above, that it is indeed meaningful to deep-learn the landscape data and that our visual representation of geometrical configurations is an efficient methodology.

3. Discussion

There are many questions in theoretical physics, or even in pure mathematics, for which one would only desire a qualitative, approximate, or partial answer, and whose full solution would often either be beyond the current scope, conceptual or computational, or would have taken considerable effort to attain. Typical such questions could be "what is the likelihood of finding a universe with three generations of particles within the landscape of string vacua or inflationary scenarios", or "what percentage of known Calabi–Yau manifolds has Hodge numbers within a prescribed range"? Attempting to address these profound questions have, with the ever-increasing power of computers, engendered our community's version of "big data", which though perhaps humble compared to some other fields, do comprise, especially considering the abstract nature of the problems at hand, of significant information often resulting from intense dialogue between teams of physicists and mathematicians for many years.

On the still-ripening fruits of this labour the philosophy of the last decade or so, particularly for the string phenomenology and computational geometry community, has been to (I) create larger and larger datasets and (II) scan through them to test the likelihood of certain salient features. Now that the data is augmenting in size and availability, it is only natural to follow the standard procedures of the data-mining community. In this letter, we have proposed the paradigm of applying deep-learning, via neural networks, such data. The purpose is twofold, the neural network can act as

Classifiers: by association of input configuration with a requisite quantity, and pattern-match over a given dataset;
Predictors: by extrapolating to hithertofore unencountered configurations, having deep-learnt a given (partial) dataset.

This is, of course, the archetypal means by which Google deep-learns the internet and hand-writing recognition software adapts to the reader's esoteric script.

It is intriguing that by going through a wealth of concrete examples from what we have dubbed **landscape data**, some of whose creation the author had been a part, this philosophy remains enlightening. Specifically, we have taken test cases from a range of problems in mathematical physics, algebraic geometry and representation theory, such as CY datasets, classification of stable vector bundles, and catalogues of quiver varieties and brane tilings. We subsequently saw that even relatively simple NN can deep-learning to extraordinary accuracy.

In some sense, this is not surprising, there is underlying structure to any classification problem in our context, which may not be manifest. Indeed, what is novel is to look at the likes of a CICY or a quiver theory as a *pixelated image*, no different from a handwritten digit, for whose analysis machine-learning has become the de facto method and a blossoming industry. *The landscape data, be they work of human hands, elements of Nature or conceptions of Mathematics, have inherent structure, sometimes more efficiently uncovered by AI via deep-learning.* Thereby, one can rapidly obtain results, before embarking on finding a reductionist framework for a fundamental theory explaining the results or proceed to intensive computations from first principles. This paradigm is especially useful when classification problems become intractable, which is often the case, here a pragmatic approach would be to deep-learn partial classification results and predict future outcome.

Under this rubric, the possibilities are endless. Several immediate and pertinent directions spring to mind. First, the largest dataset in algebraic geometry/string theory is the Kreuzer–Skarke list [5,20,21] of reflexive polytopes in dimension 4 from each of which many CY manifolds (compact and non-compact) can be constructed. To discover hidden patterns is an ongoing enterprise [14, 17] and the help of deep-learning would be a most welcome one. Next, the issue of bundle stability and cohomology is a central problem in heterotic phenomenology as well as algebraic geometry. In many ways, this is a perfect problem for machine-learning: the input is usually encodable into an integer matrix or a list of matrices, representing the coefficients in an expansion into effective divisor classes, the output is simply a vector of integers (in the case of cohomology) or a binary answer (with respective to a given Kahler class, the bundle is either stable or not). The brute-force way involves the usual spectral sequences and determining all coboundary maps or finding the lattices of subsheafs, expensive by any standards. In the case of stability checking, this is an enormous effort to arrive at a yes/no query. With increasing number of explicitly known examples of stable bundles constructed from first principles, to deep-learn this and then estimate the probability of a given bundle being stable would be tremendous time-saver.

To give an idea the high non-triviality of our venture, suppose we wanted to know how many CY3 can be constructed from the famous Kreuzer–Skarke (KS) list of 473 million reflexive polytopes. Only recently [21] was a systematic triangulation carried out on a cluster, up to $h^{1,1} = 7$ (above which even the state-of-art com-

puter is powerless), and $\simeq 100,000$ manifolds were found from $\simeq 25,000$ polytopes. The KS list has $h^{1,1}$ going up to 496, thus we have not even touched the tip of the iceberg in answering the simplest question of enumerating CY3s. Here, the NN would be extremely useful in predicting an estimate, having learnt the data from [21], which already took ~ 5000 core-hours with traditional methods on the cluster; this is currently under investigation.

We hope the reader has been persuaded by not only the scope but also the feasibility of our proposed paradigm, a paradigm of increasing importance in an Age where even the most abstruse of mathematics or the most theoretical of physics cannot avoid compilations of and investigations on perpetually growing datasets. The case studies of deep-learning such landscape of data here presented are but a few nuggets in an unfathomably vast gold-mine, rich with new science yet to be discovered.

Acknowledgements

We are indebted to the Science and Technology Facilities Council, UK, for grant ST/J00037X/1, the Chinese Ministry of Education, for a ChangJiang Chair Professorship at NanKai University, and the city of Tian-Jin for a Qian-Ren Award, as well as Merton College, University of Oxford for continued support.

References

[1] P. Candelas, G.T. Horowitz, A. Strominger, E. Witten, Vacuum configurations for superstrings, Nucl. Phys. B 258 (1985) 46.
[2] P. Candelas, A.M. Dale, C.A. Lutken, R. Schimmrigk, Complete intersection CY manifolds, Nucl. Phys. B 298 (1988) 493;
P. Candelas, C.A. Lutken, R. Schimmrigk, Complete intersection Calabi–Yau manifolds. 2. Three generation manifolds, Nucl. Phys. B 306 (1988) 113;
M. Gagnon, Q. Ho-Kim, An exhaustive list of complete intersection Calabi–Yau manifolds, Mod. Phys. Lett. A 9 (1994) 2235;
T. Hubsch, Calabi–Yau Manifolds: a Bestiary for Physicists, World Scientific, ISBN 981021927X, 1992.
[3] P. Candelas, M. Lynker, R. Schimmrigk, CY manifolds in weighted P(4), Nucl. Phys. B 341 (1990) 383.
[4] A. Grassi, D.R. Morrison, Group representations and the Euler characteristic of elliptically fibered Calabi–Yau threefolds, arXiv:math/0005196 [math-ag];
R.Y. Donagi, Principal bundles on elliptic fibrations, Asian J. Math. 1 (1997) 214, arXiv:alg-geom/9702002.
[5] A.C. Avram, M. Kreuzer, M. Mandelberg, H. Skarke, The web of Calabi–Yau hypersurfaces in toric varieties, Nucl. Phys. B 505 (1997) 625, arXiv:hep-th/9703003;
Victor V. Batyrev, Lev A. Borisov, On Calabi–Yau complete intersections in toric varieties, arXiv:alg-geom/9412017;
M. Kreuzer, H. Skarke, Reflexive polyhedra, weights and toric Calabi–Yau fibrations, Rev. Math. Phys. 14 (2002) 343, arXiv:math/0001106 [math-ag].
[6] S. Kachru, R. Kallosh, A.D. Linde, S.P. Trivedi, De Sitter vacua in string theory, Phys. Rev. D 68 (2003) 046005, arXiv:hep-th/0301240.
[7] C. Vafa, The string landscape and the swampland, arXiv:hep-th/0509212.
[8] F. Gmeiner, R. Blumenhagen, G. Honecker, D. Lust, T. Weigand, One in a billion: MSSM-like D-brane statistics, J. High Energy Phys. 0601 (2006) 004, arXiv:hep-th/0510170.
[9] V. Braun, Y.-H. He, B.A. Ovrut, T. Pantev, The exact MSSM spectrum from string theory, J. High Energy Phys. 0605 (2006) 043, arXiv:hep-th/0512177.
[10] P. Candelas, X. de la Ossa, Y.-H. He, B. Szendroi, Triadophilia: a special corner in the landscape, Adv. Theor. Math. Phys. 12 (2008) 429, arXiv:0706.3134.
[11] Y.H. He, Calabi–Yau geometries: algorithms, databases, and physics, Int. J. Mod. Phys. A 28 (2013) 1330032, arXiv:1308.0186 [hep-th].
[12] Y.-H. He, P. Candelas, A. Hanany, A. Lukas, B. Ovrut, Computational algebraic geometry in string, Gauge theory. Special issue, Adv. High Energy Phys. 2012 (2012) 431898, http://dx.doi.org/10.1155/2012/431898, Hindawi publishing.
[13] V. Braun, Toric elliptic fibrations and F-theory compactifications, J. High Energy Phys. 1301 (2013) 016, arXiv:1110.4883.
[14] W. Taylor, On the Hodge structure of elliptically fibered Calabi–Yau threefolds, J. High Energy Phys. 1208 (2012) 032, arXiv:1205.0952 [hep-th].
[15] A.P. Braun, J. Knapp, E. Scheidegger, H. Skarke, N.O. Walliser, PALP – a user manual, arXiv:1205.4147 [math.AG].
[16] P. Candelas, A. Constantin, H. Skarke, An abundance of K3 fibrations from polyhedra with interchangeable parts, arXiv:1207.4792 [hep-th].
[17] Y.H. He, V. Jejjala, L. Pontiggia, Patterns in Calabi–Yau distributions, arXiv:1512.01579 [hep-th].
[18] P. Candelas, R. Davies, New Calabi–Yau manifolds with small Hodge numbers, Fortschr. Phys. 58 (2010) 383, arXiv:0809.4681 [hep-th].
[19] W.A. Stein, et al., http://www.sagemath.org.
[20] Calabi Yau data, hep.itp.tuwien.ac.at/~kreuzer/CY/.
[21] R. Altman, J. Gray, Y.H. He, V. Jejjala, B.D. Nelson, A Calabi–Yau database: threefolds constructed from the Kreuzer–Skarke list, J. High Energy Phys. 1502 (2015) 158, arXiv:1411.1418, CY Database: www.rossealtman.com.
[22] Y.H. He, S.J. Lee, A. Lukas, Heterotic models from vector bundles on toric Calabi–Yau manifolds, J. High Energy Phys. 1005 (2010) 071, arXiv:0911.0865 [hep-th];
Y.H. He, M. Kreuzer, S.J. Lee, A. Lukas, Heterotic bundles on Calabi–Yau manifolds with small Picard number, J. High Energy Phys. 1112 (2011) 039, arXiv:1108.1031.
[23] L.B. Anderson, F. Apruzzi, X. Gao, J. Gray, S.J. Lee, A new construction of Calabi–Yau manifolds: generalized CICYs, Nucl. Phys. B 906 (2016) 441, arXiv:1507.03235 [hep-th].
[24] L.B. Anderson, Y.H. He, A. Lukas, Heterotic compactification, an algorithmic approach, J. High Energy Phys. 0707 (2007) 049, arXiv:hep-th/0702210 [hep-th].
[25] L.B. Anderson, J. Gray, A. Lukas, E. Palti, Heterotic line bundle standard models, J. High Energy Phys. 1206 (2012) 113, arXiv:1202.1757 [hep-th].
[26] L.B. Anderson, A. Constantin, J. Gray, A. Lukas, E. Palti, A comprehensive scan for heterotic SU(5) GUT models, J. High Energy Phys. 1401 (2014) 047, arXiv:1307.4787.
[27] J. Gray, A.S. Haupt, A. Lukas, All complete intersection Calabi–Yau four-folds, J. High Energy Phys. 1307 (2013) 070, arXiv:1303.1832 [hep-th].
[28] P. Gao, Y.H. He, S.T. Yau, Extremal bundles on Calabi–Yau threefolds, Commun. Math. Phys. 336 (3) (2015) 1167, arXiv:1403.1268 [hep-th].
[29] D.R. Morrison, C. Vafa, Compactifications of F theory on Calabi–Yau threefolds. 1 & 2, Nucl. Phys. B 473 (1996) 74, arXiv:hep-th/9602114, Nucl. Phys. B 476 (1996) 437, arXiv:hep-th/9603161.
[30] R. Friedman, J. Morgan, E. Witten, Vector bundles and F theory, Commun. Math. Phys. 187 (1997) 679, arXiv:hep-th/9701162.
[31] M. Gabella, Y.H. He, A. Lukas, An abundance of heterotic vacua, J. High Energy Phys. 0812 (2008) 027, arXiv:0808.2142 [hep-th].
[32] M. Cvetič, J. Halverson, D. Klevers, P. Song, On finiteness of type IIB compactifications: magnetized branes on elliptic Calabi–Yau threefolds, J. High Energy Phys. 1406 (2014) 138, arXiv:1403.4943 [hep-th].
[33] A. Hanany, Y.H. He, Nonabelian finite gauge theories, J. High Energy Phys. 9902 (1999) 013, arXiv:hep-th/9811183;
A. Hanany, Y.H. He, A monograph on the classification of the discrete subgroups of SU(4), J. High Energy Phys. 0102 (2001) 027, arXiv:hep-th/9905212.
[34] B. Feng, A. Hanany, Y.H. He, D-brane gauge theories from toric singularities and toric duality, Nucl. Phys. B 595 (2001) 165, arXiv:hep-th/0003085.
[35] S. Franco, A. Hanany, D. Martelli, J. Sparks, D. Vegh, B. Wecht, Gauge theories from toric geometry and brane tilings, J. High Energy Phys. 0601 (2006) 128, arXiv:hep-th/0505211.
[36] J. Davey, A. Hanany, J. Pasukonis, On the classification of brane tilings, J. High Energy Phys. 1001 (2010) 078, arXiv:0909.2868 [hep-th].
[37] S. Franco, Y.H. He, C. Sun, Y. Xiao, A comprehensive survey of brane tilings, arXiv:1702.03958 [hep-th].
[38] J. Davey, A. Hanany, N. Mekareeya, G. Torri, M2-branes and Fano 3-folds, J. Phys. A 44 (2011) 405401, arXiv:1103.0553 [hep-th].
[39] A. Hanany, R.K. Seong, Brane tilings and reflexive polygons, Fortschr. Phys. 60 (2012) 695, arXiv:1201.2614 [hep-th].
[40] Y.H. He, R.K. Seong, S.T. Yau, Calabi–Yau volumes and reflexive polytopes, arXiv:1704.03462.
[41] S. Franco, S. Lee, R.K. Seong, C. Vafa, Brane brick models in the mirror, J. High Energy Phys. 1702 (2017) 106, arXiv:1609.01723.
[42] Y.H. He, Deep-learning the landscape, arXiv:1706.02714.